

Extremal Results Regarding Right Angles in Vector Spaces Over Finite Fields

Michael Bennett

ABSTRACT. Here we examine some Erdős-Falconer-type problems in vector spaces over finite fields involving right angles. Our main goals are to show that

- a) a subset $A \subset \mathbb{F}_q^d$ of size $\gg q^{\frac{d+2}{3}}$ contains three points which generate a right angle, and
- b) a subset $A \subset \mathbb{F}_q^d$ of size $\gg q^{\frac{d+2}{2}}$ contains two points which generate a right angle with the vertex at the origin.

We will also prove that b) is sharp up to constants and provide some partial results for similar problems related to spread and collinear triples.

CONTENTS

1. Introduction	1
2. Main Theorems	2
3. Comparison with Euclidean Analog	2
4. Proof of Theorem 2.2	3
5. Proof of Theorem 2.4	5
6. Spread	9
7. Collinear triples	12
References	14

1. Introduction

In this paper, we will be looking at some extremal problems in combinatorial geometry similar to those proposed by Erdős and Falconer, but in the setting of finite fields rather than in Euclidean space. Throughout, q will be a power of an odd prime, and we will denote the field with q elements by \mathbb{F}_q . We would like to determine how large a subset of \mathbb{F}_q^d (with q much larger than d) needs to be to guarantee that the set contains three points which form a right angle. We will also make some observations about extremal problems regarding “spread.” Spread is essentially a finite field analogue of the Euclidean angle. When we say that a triple (a, b, c) “generates” an angle (or spread) θ , we mean that the angle between the vectors $a - b$ and $c - b$ is θ . This is, in a way, an extension of the right angle problem, because to say that a triple “generates a right angle” is roughly equivalent to saying that the spread generated by the triple is 1. These notions will be explained much more

thoroughly in section 6. Finally, we will ask how large a subset of \mathbb{F}_q^d needs to be to guarantee that three points in that subset are collinear and provide relatively trivial bounds on this problem, not coming particularly close to a decisive answer to the question. This paper was motivated by a paper of Harangi et. al. which looked at the same problems, as well as many others, in \mathbb{R}^d . By looking at the same problems in a different setting, we hope to improve our understanding of the relation between the geometric structure of \mathbb{R}^d and \mathbb{F}_q^d .

2. Main Theorems

Throughout, we will use the \gg symbol. We say that $X \gg Y$ if $X = \Omega(Y)$ with respect to some variable (often q or p).

DEFINITION 2.1. We say that an ordered triple of points $(x, y, z) \in \mathbb{F}_q^d \times \mathbb{F}_q^d \times \mathbb{F}_q^d$ forms a right angle if x, y , and z are distinct, and the vectors $x - y$ and $z - y$ have dot product 0.

THEOREM 2.2. *If $A, B \subset \mathbb{F}_q^d$, $A \cap B = \emptyset$, and $|A|^2|B| \gg q^{d+2}$, then there are $x, y \in A$ and $z \in B$ so that (x, y, z) forms a right angle.*

Following immediately from the theorem, we have:

COROLLARY 2.3. *If $A \subset \mathbb{F}_q^d$ and $|A| \gg q^{\frac{d+2}{3}}$ then A contains a right angle.*

That is, given a set A , we may decompose it into disjoint sets A' and B' of roughly equal size and then apply the theorem. We will see that the bound from 2.2 is best possible up to the implied constant. It is not known whether the result is improvable in the form of 2.3.

We will also examine the problem of existence of right angles whose vertex is fixed at the origin:

THEOREM 2.4. *If $A \subset \mathbb{F}_q^d$ and $|A| \gg q^{\frac{d+2}{2}}$, then there are $x, y \in A$ so that $x \cdot y = 0$. Moreover, if q is a prime, then there exists a subset $B \subset \mathbb{F}_q^d$ so that $|B| \gg q^{\frac{d+2}{2}}$ but $\{x, y \in B : x \cdot y = 0\} = \emptyset$.*

3. Comparison with Euclidean Analog

Perhaps the most interesting aspect of these results are how they differ from those found in [4]. In their paper, Harangi et al. solve Euclidean versions of the problems we are solving here. There they pose the problem in two different ways: For a given angle, how large must a set in \mathbb{R}^d be to guarantee that

- 1) that angle is generated by that set?
- 2) an angle with size within δ of the given angle is generated by that set?

When we say “how large,” we are referring to the Hausdorff dimension of the set. In the paper they discover that 90° angles are a singular case. In the sense of question 1, the critical dimension for 90° angles is between $\frac{d}{2}$ and $\frac{d+1}{2}$. That is, there are examples of sets of Hausdorff dimension arbitrarily close to $\frac{d}{2}$ which contain no 90° angles, while any compact set with Hausdorff dimension larger than $\frac{d+1}{2}$ must contain a 90° angle (which they prove rather concisely in the paper). For 0 and 180° , the critical dimension is $d - 1$ (because of the $(d - 1)$ -sphere), while for other angles only partial answers exist. Question 2 is completely resolved in this paper, where they find that a set of

Hausdorff dimension 1 always has an angle very close to 90° , while a set of lower dimension need not. It is known that any sufficiently large finite set of points contains three points at an angle as close to 180° as desired (and therefore close to 0° as well) (see [3]). Interestingly, 60° and 120° angles also have a separate result that depends on δ . For any other given angle, however, a set of Hausdorff dimension that increases to ∞ as d goes to ∞ can be constructed so that a neighborhood of that angle is avoided.

In Euclidean space, Hausdorff dimension is a natural way to classify the dimension of an arbitrary set. The best analog to this classification in vector spaces over finite fields is to think of the dimension of a set S as $\approx \log_q(|S|)$. This is why, in the finite field setting, the quantity we are most interested in is the exponent α in $|S| = cq^\alpha$. In extremal geometry, problems that can be solved in Euclidean space can often be solved in a similar manner over finite fields. Consequently, one might expect the exponent in the finite field setting to match the Hausdorff dimension in the Euclidean setting. This is why the case of right angles is interesting here. We find that the critical dimension in vector spaces over finite fields is bounded above by $\frac{d+2}{3}$, which is significantly different from the (question 1) threshold found in [4]. As far as the Euclidean analog of theorem 2.2, consider the set $(0, \infty) \times \dots \times (0, \infty) \subset \mathbb{R}^d$. The dot product of any two elements in this set must be positive, and the set clearly has Hausdorff dimension d , so this also presents a discrepancy between the continuous and finite settings.

4. Proof of Theorem 2.2

We begin with the following lemma.

LEMMA 4.1. *Let H be a nonempty set of hyperplanes in \mathbb{F}_q^d . Let $E = \bigcup_{h \in H} h$. Then $|E^c| \leq \frac{q^{d+1}}{|H|}$.*

PROOF. We will use ‘ h ’ as the characteristic function for the set h . First we note that

$$(4.1) \quad \sum_{h \in H, x \in \mathbb{F}_q^d} h(x) = \sum_{h \in H} |h| = q^{d-1}|H|.$$

Using Cauchy-Schwarz and the fact that every $h \in H$ is a subset of E , we also have

$$\begin{aligned} \sum_{h \in H, x \in \mathbb{F}_q^d} h(x) &= \sum_{h \in H, x \in E} h(x) \leq \left(\sum_{x \in E} 1^2 \right)^{1/2} \left(\sum_{x \in E} \left(\sum_{h \in H} h(x) \right)^2 \right)^{1/2} \\ &= |E|^{1/2} \left(\sum_{h_1, h_2 \in H} \sum_{x \in \mathbb{F}_q^d} h_1(x) h_2(x) \right)^{1/2}. \end{aligned}$$

If $h_1 = h_2$, then $\sum_x h_1(x) h_2(x) = q^{d-1}$, and otherwise $\sum_x h_1(x) h_2(x) \leq q^{d-2}$. Making use of equation 4.1, we now have

$$q^{d-1}|H| \leq |E|^{1/2} (q^{d-1}|H| + q^{d-2}|H|^2)^{1/2}.$$

Solving for $|E|$ gives

$$|E| \geq \frac{q^{2d-2}|H|^2}{q^{d-1}|H| + q^{d-2}|H|^2} = \frac{q^d|H|}{q + |H|}.$$

Lastly,

$$|E^c| = q^d - |E| \leq \frac{q^{d+1}}{q + |H|} \leq \frac{q^{d+1}}{|H|}.$$

□

We are now ready to prove the theorem. For any ordered pair $(x, y) \in \mathbb{F}_q^d \times \mathbb{F}_q^d$ with $x \neq y$, Define h_{xy} to be

$$\{z : (z - x) \cdot (y - x) = 0\},$$

i.e. the (translated) hyperplane through the point x with normal vector in the direction of $y - x$.

Let $A, B \subset \mathbb{F}_q^d$ with $A \cap B = \emptyset$. Let A' be any subset of A with $|A'| = \left\lfloor \frac{|A|}{2} \right\rfloor$. We consider the set of hyperplanes $H = \{h_{xy} : (x, y) \in A' \times B, x \neq y\}$. Notice first, that if $c \in (h_{ab} \setminus \{a\}) \cap A$, then (b, a, c) defines a right angle. Thus we will assume that $h_{ab} \cap A = \{a\}$ for all $a \in A, b \in B$. Consequently, $h_{ab} = h_{cd}$ only when $a = c$ (as otherwise $h_{ab} \cap A$ contains two distinct elements a and c). It is also important to note that if $\|x - y\| = 0$, then $y \in h_{xy}$. We want to ensure that x is the only point of A in h_{xy} , which is why we insist that A and B be disjoint. Finally, notice that for fixed a and b , there are exactly $q - 1$ choices of b' so that $h_{ab} = h_{ab'}$, i.e. all points on the line through a and b except for a .

Because there are $|A'| |B|$ elements in $A' \times B$, there must be at least $\frac{|A'| |B|}{q-1}$ distinct planes in H .

Let $E = \bigcup_{h \in H} h$. By assumption, $A \setminus A' \subset E^c$. Notice that $|A \setminus A'| = \left\lceil \frac{|A|}{2} \right\rceil$. Using Lemma 4.1, we have

$$|A \setminus A'| \leq \frac{q^{d+1}(q-1)}{|A'| |B|}$$

or equivalently,

$$\left\lceil \frac{|A|}{2} \right\rceil \leq \frac{q^{d+1}(q-1)}{\left\lfloor \frac{|A|}{2} \right\rfloor |B|}$$

Solving for $|A|$ and $|B|$ gives us

$$|A|^2 |B| \leq 4q^{d+2}$$

In other words, if none of the points of B completes a right angle with any of the pairs of points in A , then $|A|^2 |B|$ can be no larger than $4q^{d+2}$.

4.1. Sharpness. If $d \geq 3$ then we can find an isotropic line $L \subset \mathbb{F}_q^d$ (see 6.1). Indeed, if x is an isotropic vector then the line through x and the origin has the aforementioned property.

Now let L be an isotropic line and let H be the hyperplane that is perpendicular to L . Remember, because L is isotropic, $L \subset H$. Because of this, the only hyperplane parallel to H that meets L is H itself. Then, in the language of theorem 2.2, we choose $A = L$ and $B = \mathbb{F}_q^d \setminus H$. Notice that $|A|^2 |B| = q^{d+2}(1 + o(1))$. If $x, y \in A$ and $z \in B$, then

$$(x - y) \cdot (z - y) = z \cdot (x - y) - x \cdot y + y \cdot y = z \cdot (x - y).$$

Now $x - y \in L$ and z is not orthogonal to L by construction, so $(x - y) \cdot (z - y) \neq 0$, which completes the proof. In fact, this setup also works for $d = 2$ whenever $q \equiv 1 \pmod{4}$, since \mathbb{F}_q contains a square root of -1 , and hence an isotropic line. (In dimension 2, notice that an isotropic line is its own orthogonal hyperplane).

5. Proof of Theorem 2.4

This theorem has already been proved by Hart et. al. in [5], including results for nonzero dot products as well. In fact they show that if $|A| \gg q^{\frac{d+1}{2}}$, then for any nonzero scalar t , there are two points in A whose dot product is t . Here we provide a slightly different proof here of the 0 case, using very similar techniques to those in [5], to keep this paper more self-contained. More importantly, we also show that the 0 case is an exceptional case—that a guaranteed dot product of 0 requires about $q^{\frac{d+2}{2}}$ points of \mathbb{F}_q^d , at least in the case q is prime.

We will make use of the Fourier transform here. Let $\chi : \mathbb{F}_q \rightarrow \mathbb{C}$ be a nontrivial additive character. We define the Fourier transform \hat{f} of a function $f : \mathbb{F}_q^d \rightarrow \mathbb{C}$ to be

$$\hat{f}(\xi) = \frac{1}{q^d} \sum_{\eta \in \mathbb{F}_q^d} f(\eta) \chi(-\xi \cdot \eta).$$

If we let A be its own characteristic function, then we can write

$$\#\{x, y \in A : x \cdot y = 0\} = \frac{1}{q} \sum_{x, y \in \mathbb{F}_q^d} A(x) A(y) \sum_{s \in \mathbb{F}_q} \chi(sx \cdot y).$$

Separating the cases where $s = 0$ and $s \neq 0$, we get

$$\#\{x, y \in A : x \cdot y = 0\} = \frac{|A|^2}{q} + \frac{1}{q} \sum_{x, y \in \mathbb{F}_q^d} A(x) A(y) \sum_{s \in \mathbb{F}_q^*} \chi(sx \cdot y),$$

where $\mathbb{F}_q^* = \mathbb{F}_q \setminus \{0\}$.

Using the Fourier transform, we get

$$\#\{x, y \in A : x \cdot y = 0\} = \frac{|A|^2}{q} + q^{d-1} \sum_{x \in \mathbb{F}_q^d, s \in \mathbb{F}_q^*} A(x) \hat{A}(-sx),$$

and using Cauchy-Schwarz, we can bound this expression below by

$$\frac{|A|^2}{q} - q^{d-1} \sum_{s \in \mathbb{F}_q^*} \sqrt{\sum_{x \in \mathbb{F}_q^d} |A(x)|^2} \sqrt{\sum_{x \in \mathbb{F}_q^d} |\hat{A}(sx)|^2}.$$

Applying Plancherel's theorem to the rightmost sum and noting that $\sum_{x \in \mathbb{F}_q^d} |A(x)|^2 = |A|$, we have

$$\#\{x, y \in A : x \cdot y = 0\} \geq \frac{|A|^2}{q} - q^{\frac{d}{2}} |A|.$$

Since this expression is positive if $|A| > q^{\frac{d+2}{2}}$, the theorem is proven.

5.1. Sharpness. We should first note that in the case $d = 3$, the sharpness of this result has essentially been proved already by Mubayi and Williford. In [6], they find a set of mutually non-orthogonal lines through the origin in \mathbb{F}_q^3 of cardinality $\Theta(q^{3/2})$. In fact, they construct explicit sets of lines for every $q = p^n$ (separating $p = 2$ and $p \neq 2$ cases, as well as cases of even and odd n) and find tight bounds on the sizes of these sets. Since all of these lines pass through the origin, the set of all points lying on these lines has size $\Theta(q^{5/2})$. By construction, no pair of these points can have dot product zero, and $\Theta(q^{5/2})$ is sharp per 2.4. (Mubayi and Williford also prove that their estimate is sharp using graph theoretic techniques).

Here we will show that for any d , there is a set of mutually non-orthogonal lines through the origin in \mathbb{F}_p^d of size $\Theta(p^{\frac{d}{2}})$. This will in turn give us a set of $\Theta(p^{\frac{d+2}{2}})$ points, showing that our estimate in 2.4 is sharp up to a factor depending on d . Notice, however, that we insist on a prime p , as our method of proof calls for it. Furthermore, we will not find the exact size of the set of points we use here, but a lower bound of $Cp^{\frac{d+2}{2}}$, where C is almost certainly not optimal.

Order the elements of \mathbb{F}_p in the natural way, from 0 to $p-1$. Let $d \geq 2$ (but much smaller than p) and define $\sigma = \lfloor \sqrt{p/d} \rfloor$ for ease of notation. Notice that $\lfloor \sqrt{p/d} \rfloor \neq \sqrt{p/d}$, as $d \nmid p$. Let $A = \{a \in \mathbb{F}_q : 1 \leq a \leq \sigma\}$ and let $A^d \subset \mathbb{F}_q^d$ be the d -fold cross product. Notice that if $x, y \in A^d$ then $d \leq x \cdot y \leq p-1$. Right away, we see the set A^d contains no pair of points whose dot product is 0. We will show that the number of lines through the origin that meet A^d is $\Theta(p^{\frac{d}{2}})$.

For any point $x \in A^d$, let l_x be the line through x and the origin. Then the number of lines through the origin that meet A^d is precisely

$$\sum_{x \in A^d} \frac{1}{|l_x \cap A^d|},$$

so let's compute $|l_x \cap A^d|$.

It will be easier to work in \mathbb{Z} for now and then mod out by $p\mathbb{Z}$ later.

For $j, k \in \mathbb{Z}^+$, define $S_k^j = \left\{ \left\lfloor \frac{jp}{k} \right\rfloor + i : 1 \leq i \leq \left\lfloor \frac{\sigma}{k} \right\rfloor \right\}$.

Let $S_k = \bigcup_{j=0}^{k-1} S_k^j$.

Then $S_k \subset \{s : 0 \leq s \leq p-1, 1 \leq (ks) \bmod p \leq \sigma\}$ (where $(ks) \bmod p$ really means $ks - p \left\lfloor \frac{ks}{p} \right\rfloor$).

Notice that if $y \in l_x$, then y is a scalar multiple of x . It follows that if $x = (x_1, x_2, \dots, x_d)$, then

$$\bigcap_{i=1}^d S_{x_i} \subset l_x \cap A^d.$$

Now let $m_1, m_2 \in A$ with $m_1 \leq m_2$. Then

$$S_{m_1}^{j_1} \cap S_{m_2}^{j_2} = \left\{ \left\lfloor \frac{j_1 p}{m_1} \right\rfloor + i : 1 \leq i \leq \left\lfloor \frac{\sigma}{m_1} \right\rfloor \right\} \cap \left\{ \left\lfloor \frac{j_2 p}{m_2} \right\rfloor + i : 1 \leq i \leq \left\lfloor \frac{\sigma}{m_2} \right\rfloor \right\}.$$

If $\frac{j_1 p}{m_1} = \frac{j_2 p}{m_2}$, then $S_{m_1}^{j_1} \cap S_{m_2}^{j_2} = \left\{ \left\lfloor \frac{j_2 p}{m_2} \right\rfloor + i : 1 \leq i \leq \left\lfloor \frac{\sigma}{m_2} \right\rfloor \right\}$. Otherwise,

$$\left| \frac{j_1 p}{m_1} - \frac{j_2 p}{m_2} \right| \geq \frac{p}{m_1 m_2} > \frac{\sigma}{m_1}.$$

Therefore $S_{m_1}^{j_1} \cap S_{m_2}^{j_2} = \emptyset$ unless $\frac{j_1 p}{m_1} = \frac{j_2 p}{m_2}$.

This makes it very easy to calculate $|S_{m_1} \cap S_{m_2}|$. We have

$$|S_{m_1} \cap S_{m_2}| = \left\lfloor \frac{\sigma}{m_2} \right\rfloor \# \left\{ (j_1, j_2) : \frac{j_1}{m_1} = \frac{j_2}{m_2}, 0 \leq j_1 \leq m_1 - 1, 0 \leq j_2 \leq m_2 - 1 \right\}.$$

The size of the set on the right is simply $\gcd(m_1, m_2)$. Extrapolating, it is easy to see that

$$\left| \bigcap_{i=1}^d S_{x_i} \right| = \left\lfloor \frac{\sigma}{\max_i \{x_i\}} \right\rfloor \gcd(x_1, x_2, \dots, x_d)$$

Thus, the number of lines through the origin that meet A^d is

$$\sum_{x \in A^d} \frac{1}{\left\lfloor \frac{\sigma}{\max_i \{x_i\}} \right\rfloor \gcd(x_1, x_2, \dots, x_d)} \geq \sigma^{-1} \sum_{x \in A^d} \frac{\max_i \{x_i\}}{\gcd(x_1, x_2, \dots, x_d)}.$$

We can use $\frac{\max\{x_1, x_2\}}{\gcd(x_1, x_2)}$ as a lower bound for the summand, which may seem like a significant loss of precision. However, we will see that we only lose a best-possible coefficient for our main term.

Now we have

$$\begin{aligned} & \sigma^{-1} \sum_{(x_1, x_2) \in A^2} \frac{\max\{x_1, x_2\}}{\gcd(x_1, x_2)} \sum_{1 \leq x_3, \dots, x_d \leq \sigma} 1 \\ &= \sigma^{d-3} \sum_{(x_1, x_2) \in A^2} \frac{\max\{x_1, x_2\}}{\gcd(x_1, x_2)} \\ &= -\sigma^{d-2} + 2\sigma^{d-3} \sum_{1 \leq x_1 \leq x_2 \leq \sigma} \frac{x_2}{\gcd(x_1, x_2)} \\ &= -\sigma^{d-2} + 2\sigma^{d-3} \sum_{x_2=1}^{\sigma} \sum_{e|x_2} \frac{x_2}{e} \sum_{\substack{1 \leq x_1 \leq \sigma \\ \gcd(x_1, x_2)=e}} 1 \\ &= -\sigma^{d-2} + 2\sigma^{d-3} \sum_{x_2=1}^{\sigma} \sum_{e|x_2} \frac{x_2}{e} \varphi\left(\frac{x_2}{e}\right) \\ &= -\sigma^{d-2} + 2\sigma^{d-3} \sum_{x_2=1}^{\sigma} \sum_{e|x_2} e \varphi(e) \end{aligned}$$

where φ is the Euler totient function.

Changing the order of summation, we have

$$\begin{aligned}
&= -\sigma^{d-2} + 2\sigma^{d-3} \sum_{e=1}^{\sigma} e\varphi(e) \sum_{\substack{1 \leq x_2 \leq \sigma \\ e|x_2}} 1 \\
&= -\sigma^{d-2} + 2\sigma^{d-3} \sum_{e=1}^{\sigma} \left\lfloor \frac{\sigma}{e} \right\rfloor e\varphi(e) \\
&\geq -\sigma^{d-2} + 2\sigma^{d-3} \sum_{e=1}^{\sigma} (\sigma - e)\varphi(e)
\end{aligned}$$

Now we will make use of the following estimate (see, for instance, [7]):

$$\sum_{i=1}^n \varphi(i) = \frac{3}{\pi^2} n^2 + O(n \log n).$$

Using this (along with summation by parts) we can calculate $\sum_{i=1}^n i\varphi(i)$ as well:

$$\begin{aligned}
\sum_{i=1}^n i\varphi(i) &= \sum_{i=1}^n \sum_{k=1}^i \varphi(i) = \sum_{k=1}^n \sum_{i=k}^n \varphi(i) \\
&= \sum_{k=1}^n \frac{3}{\pi^2} [(n^2 - k^2) + O(n \log n)] = (1 - o(1)) \frac{2}{\pi^2} n^3
\end{aligned}$$

where $o(1)$ stands for some a_n which goes to 0 as $n \rightarrow \infty$. Now we know that the number of lines through the origin that hit a point of A^d is at least

$$\begin{aligned}
&-\sigma^{d-2} + 2\sigma^{d-3} \sum_{1 \leq e \leq \sigma} (\sigma - e)\varphi(e) \\
&= -\sigma^{d-2} + 2\sigma^{d-3} (1 - o(1)) \left(\frac{3}{\pi^2} \sigma^3 - \frac{2}{\pi^2} \sigma^3 \right) \\
&= \frac{2}{\pi^2} (1 - o(1)) \sigma^d = (1 - o(1)) \frac{2}{\pi^2} \left(\frac{p}{d} \right)^{d/2}.
\end{aligned}$$

As discussed above, no two points of $\bigcup_{x \in A^d} l_x$ have a dot product of zero. We have just concluded that

$$\left| \bigcup_{x \in A^d} l_x \right| = p \sum_{x \in A^d} \frac{1}{|l_x \cap A^d|} \geq \frac{p^{\frac{d+2}{2}}}{5d^{\frac{d}{2}}},$$

and therefore our result from 2.4 is sharp up to constants.

6. Spread

The length of a vector $a \in \mathbb{F}_q^d$ is defined by

$$\|a\| = \sum_{i=1}^d a_i^2.$$

We will use $\|a - b\|$ to refer to the distance between points a and b . Without the square root, this definition does not agree with the Euclidean space definition, and doesn't fit the qualifications of a metric; however it does retain the important property of invariance under the action of the orthogonal group.

The spread S between two vectors $a, b \in \mathbb{F}_q^d$ is defined by

$$S(a, b) = 1 - \frac{(a \cdot b)^2}{\|a\| \|b\|}.$$

We see that this definition is somewhat consistent with that of angles in Euclidean space, as

$$\sin^2(\theta) = 1 - \frac{(a \cdot b)^2}{|a|^2 |b|^2}.$$

It might seem more natural to define spread in analogy with the $\sin \theta$ or $\cos \theta$ formula, rather than $\sin^2 \theta$. However, the square root operation is not well-defined from $\mathbb{F}_q \rightarrow \mathbb{F}_q$, so this is essentially impossible. If we were to define angles in *Euclidean* space via $\theta = \phi$ iff $\sin^2 \theta = \sin^2 \phi$, there would be no distinction between an angle and its supplement. But in vector spaces over finite fields, two such angles cannot be distinguished, because rays and lines are identical objects. That is, a ray originating at the origin through the point b also goes through the point $-b$. So spread is a perfectly reasonable analog to the Euclidean angle.

DEFINITION 6.1. An isotropic vector is a vector \vec{v} such that $\vec{v} \neq 0$ but $\|\vec{v}\| = 0$. An isotropic line is a line through the origin in the direction of an isotropic vector.

Notice that for any two vectors a, b on an isotropic line, $a \cdot b = 0$.

Our definition of spread is a bit problematic with isotropic vectors since we need to divide by vector length to determine the spread. We will just say the spread is undefined in this situation. Isotropic vectors can be found in \mathbb{F}_q^d for $d > 2$ and in \mathbb{F}_q^2 when $q \equiv 1 \pmod{4}$. Fortunately, in the arguments we present here, isotropic vectors are not a significant obstacle.

Here are a few easily verified properties of spread:

$$S(a, b) = S(ra, sb) \text{ for any } r, s \in \mathbb{F}_q^*.$$

$$S(a, b) = S(b, a).$$

$$S(a, b) = S(\sigma(a), \sigma(b)) \text{ where } \sigma \text{ is an element of the orthogonal group.}$$

These are consistent with properties of the angle between vectors in Euclidean space. Notice that two vectors with a spread of 1 have a dot product of 0, and the converse is true for non-isotropic vectors. Therefore, we can think of two vectors that have a spread of 1 as making a right angle. In the next theorem we look at other spreads, which have surprisingly different results from spread 1.

We have the following theorem from [5]:

THEOREM 6.2. *Suppose $E \subset S_t^{d-1} \subset \mathbb{F}_q^d$ and $|E| \gg q^{\frac{d}{2}}$. Then the points of E generate all distances if d is even and generate a positive proportion of distances if d is odd.*

We will use this result to make the proof of our next theorem rather easy.

THEOREM 6.3. *Let $d \geq 3$, $E \subset \mathbb{F}_q^d$, $|E| \geq Cq^{\frac{d+2}{2}}$ for some sufficiently large constant C . Then there is a constant c independent of q so that at least cq distinct spreads occur in E .*

There is no evidence that the exponent $\frac{d+2}{2}$ is best-possible. It can likely be improved using more sophisticated methods.

It is also worth mentioning a related result of Vinh ([8]): Given a set $E \subset S_t^2$ of size $q^{\frac{d+3}{2}}$,

$$\#\{(a, b) \in S_t^2 \times S_t^2 : S(a, b) = s\} = O\left(\frac{|E|^2}{q}\right)$$

when $1 - s$ is a square in \mathbb{F}_q .

REMARK 6.4. Notice that if $a, b \in \mathbb{F}_q^d$ and $\|a\| = \|b\|$, then $1 - S(a, b) = 0$ or is a square in \mathbb{F}_q . Therefore, if S^{d-1} is the $(d-1)$ -dimensional sphere of radius t in \mathbb{F}_q^d , then $\{1 - S(a, b) : a, b \in S_t^{d-1}\}$ contains no nonsquares. If we were to look at the set of spreads generated by triples of points on S_t^{d-1} , we may very well get all spreads. However, in our proof of the above theorem, our strategy will be to look at spreads of triples $(a, \vec{0}, b)$, and thus we will only guarantee a positive proportion of spreads.

PROOF. Let $E \subset \mathbb{F}_q^d$ with $|E| > C_1 q^{\frac{d+2}{2}}$, and suppose, without loss of generality, that $\vec{0} \in E$. With C_1 sufficiently large, there must be some $t \in \mathbb{F}_q^*$ so that the size of the sphere S_t^{d-1} of radius t centered at the origin has size $> C_2 q^{\frac{d}{2}}$ by pigeonholing. For simplicity, we will suppose $t = 1$, but the following arguments work for any fixed value of t .

Next we'll show that for points a, b, c, d on the sphere,

$$S(a, b) = S(c, d) \text{ if and only if } \|a - b\| = \|c \pm d\|.$$

First, notice that $\|a - b\| = 2 - 2a \cdot b$. If $\|a - b\| = \|c - d\|$, then $a \cdot b = c \cdot d$. If $\|a - b\| = \|c - (-d)\|$, then $a \cdot b = -c \cdot d$. Either way, we have $(a \cdot b)^2 = (c \cdot d)^2$, and thus $S(a, b) = S(c, d)$. This argument can be reversed to recover the other direction.

We now know that, given a positive proportion of the distances on the sphere, we generate a positive proportion of the spreads on the sphere. Invoking 6.2, we know that if $|E \cap S_t^{d-1}| \gg q^{\frac{d}{2}}$ then the set of triples $\{(a, \vec{0}, b) : a, b \in E \cap S_t^{d-1}\}$ generate a positive proportion of spreads. \square

This result says nothing of substance about the 2-dimensional case. We will instead prove the following:

THEOREM 6.5. *If $A \subset \mathbb{F}_q^2$ and $|A| \geq 2q - 1$ then A generates all spreads.*

This is, of course, sharp up to the coefficient of q , as a line in \mathbb{F}_q^2 has q points but only generates one spread (namely spread = 0, as long as the line is not parallel to an isotropic line).

Before we start the proof, we need to make an observation. The “spread between two lines,” where we think of the vertex as the intersection point of the lines, is well-defined because

1) $S(a, b) = S(ra, sb)$ for any $r, s \in \mathbb{F}_q^*$ (as stated earlier)

and

2) the spread generated by the triple (a, b, c) is the same as the spread generated by $(a+z, b+z, c+z)$ for any $z \in \mathbb{F}_q^2$.

Property 1 shows that if b is the intersection of two lines, then the spread generated by (a, b, c) will be consistent for any choice of $a \neq b$ on the first line, and any choice of $c \neq b$ on the second. Property 2 shows that the point of intersection need not be the origin. Moreover, it is easy to see that if l_1 and l_2 are parallel lines, then for any other nonparallel line l_0 (as long as it is not a translation of an isotropic line), the spread between l_1 and l_0 is equal to the spread between l_2 and l_0 . Now we may continue with the proof.

PROOF. We will let $A \subset \mathbb{F}_q^2$ so that $|A| = 2q - 1$. If A is larger, we may simply choose a subset of size $2q - 1$. Choose any spread $\theta \neq 0$ and let μ, ν be two different lines through the origin with spread θ between them. Let L_μ be the set of lines parallel to μ and L_ν the set of lines parallel to ν . Clearly, $|L_\mu| = |L_\nu| = q$. It suffices to show that there are three points $a, b, c \in A$ so that b and a lie on the same line in L_μ and b and c lie on the same line of L_ν . Notice, by pigeonholing, that

$$\#\{x \in A : |l_\mu(x) \cap A| \geq 2\} \geq q.$$

where $l_\mu(x)$ is the line in L_μ containing the point x . In other words, there are at least q points in A that share a line of L_μ with another point of A . This of course applies to lines of L_ν also. By further pigeonholing, since there are only $2q - 1$ points to begin with, there must be a point b of A which shares its L_μ line with a point $a \in A$ and its L_ν line with a point $c \in A$. Thus (a, b, c) produces spread θ in A . \square

REMARK 6.6. Unfortunately, we had to avoid spread 0 in the above argument. In dimension 2, three points generate a spread of 0 only when they are collinear. If we try to apply the above argument to this case, then we get $L_\mu = L_\nu$ and the argument breaks down. We can still show that $2q - 1$ points in \mathbb{F}_q^2 are enough to produce a spread of 0, but we will do this in the next section using a different argument. Before doing so, we should note that there is an important difference between 180° angles in Euclidean space and “ 180° angles” (i.e. spread 0) in vector spaces over finite fields. One can easily check that, in \mathbb{F}_q^d , three collinear points give a spread of 0 (or an undefined spread in the case that they lie on a line parallel to an isotropic line). However, in dimensions higher than 2, three points that generate a spread of 0 need not be collinear. Take, for example, the triple $((0, 0, 0), (1, 0, 0), (2, 1, 2)) \in \mathbb{F}_5^3 \times \mathbb{F}_5^3 \times \mathbb{F}_5^3$.

7. Collinear triples

We want to know how large a subset of \mathbb{F}_q^d must be to guarantee that it contains three collinear points. (See [1] for an examination of this problem in \mathbb{F}_3^d .) We will not resolve this problem here, but we will give some partial results.

7.1. Upper Bound. First, it is rather simple to show that any set E containing more than $\frac{q^d-1}{q-1} + 1$ points must have three collinear points:

Suppose E has $\frac{q^d-1}{q-1} + 2$ points and choose any point $p_0 \in E$. Then there are $\frac{q^d-1}{q-1}$ lines of \mathbb{F}_q^d running through this point. Counting p_0 and allowing for one additional point on each line through p_0 gives us a total of $\frac{q^d-1}{q-1} + 1$ points. Since E contains more points than this, one of the lines through p_0 must contain at least two other points of E .

In the case that d is even, we can improve this result ever-so-slightly. Suppose that there is a set E of $\frac{q^d-1}{q-1} + 1$ points with no more than two on any line. Since there are $\frac{q^d-1}{q-1}$ lines through each point, every line through a point of E must go through two points of E . Choose any line l in \mathbb{F}_q^d and let L be the set of lines parallel to l . Every point in \mathbb{F}_q^d is contained in some line of L and therefore the number of lines of L that meet E must be exactly $\frac{1}{2} \left(\frac{q^d-1}{q-1} + 1 \right)$. However, if d is even (and recall that we assume q is odd throughout this paper), then this quantity is not an integer, giving a contradiction. We conclude that any subset with more than $\frac{q^d-1}{q-1}$ points must have three on a line.

Obviously this is not a significant improvement, but it is the best possible bound in the case $d = 2$. Consider the set

$$\{(x, y, t) \in (\mathbb{F}_q \setminus \{-1, 1\}) \times \mathbb{F}_q^* \times \mathbb{F}_q^* : x^2 - ty^2 = 1\}.$$

Since each choice of (x, y) determines a unique t , this set has cardinality $(q-1)(q-2)$. Therefore, by pigeonholing, there is some $t \in \mathbb{F}_q^*$ so that

$$\#\{(x, y) \in (\mathbb{F}_q \setminus \{-1, 1\}) \times \mathbb{F}_q^* : x^2 - ty^2 = 1\} \geq q - 2.$$

Since $(1, 0)$ and $(-1, 0)$ are also solutions to $x^2 - ty^2 = 1$, we see that for this t ,

$$\#\{(x, y) \in \mathbb{F}_q^2 : x^2 - ty^2 = 1\} \geq q.$$

Finally, since solutions to this equation always come in pairs (i.e. $a^2 - tb^2 = 1$ implies $(-a)^2 - t(-b)^2 = 1$), the total number of solutions must be at least $q + 1$, since q is odd.

Let $E = \{(x, y) \in \mathbb{F}_q^2 : x^2 - ty^2 = 1\}$. We will show now that no three points in E are collinear. Fix any nonzero t and let $y = c$ be a fixed horizontal line in \mathbb{F}_q^2 . The points of intersection with E are those (x, c) satisfying $x^2 - tc^2 = 1$. Since there are at most two values of x solving this equation, any horizontal line meets E in at most two points.

Let $x = my + n$ be any non-horizontal line in \mathbb{F}_q^2 . Again, there are at most two values of y satisfying

$$(my + n)^2 - ty^2 = 1$$

and the value of x is determined uniquely by the value of y because of our line equation. Therefore, no line hits E in more than two points.

7.2. Lower Bound. It is tempting to guess that a $(d-1)$ -sphere in \mathbb{F}_q^d should only grant two points per line and would make our above observation sharp. Alas, this is not the case, as spheres of dimension greater than 2 always contain lines (in the finite field setting). In fact, any hypersurface in $(\overline{\mathbb{F}}_q)^d$ ($\overline{\mathbb{F}}_q$ is the algebraic closure of \mathbb{F}_q) with degree less than $2d-3$ contains several entire lines (see [2]). However, for any d we can provide a set P of size $q^{\lfloor \frac{2d}{3} \rfloor}$ for which any line in \mathbb{F}_q^d contains at most two points of P . This set is, in essence, a product of graphs of 2-dimensional paraboloids. We let t be any nonsquare in \mathbb{F}_q and then let

$$P_d = \begin{cases} \bigoplus_{i=1}^{\frac{d}{3}} \{(x, y, tx^2 - y^2) : x, y \in \mathbb{F}_q\}, & d \equiv 0 \pmod{3} \\ \left(\bigoplus_{i=1}^{\frac{d-1}{3}} \{(x, y, tx^2 - y^2) : x, y \in \mathbb{F}_q\} \right) \oplus \{0\}, & d \equiv 1 \pmod{3} \\ \left(\bigoplus_{i=1}^{\frac{d-2}{3}} \{(x, y, tx^2 - y^2) : x, y \in \mathbb{F}_q\} \right) \oplus \{(x, tx^2) : x \in \mathbb{F}_q\}, & d \equiv 2 \pmod{3} \end{cases}$$

One can see that, for any choice of d , $|P_d| = q^{\lfloor \frac{2d}{3} \rfloor}$.

To prove that P_d does not contain 3 collinear points, we first show that any line in \mathbb{F}_q^3 has at most two points in common with $P_3 = \{(x, y, tx^2 - y^2) : x, y \in \mathbb{F}_q\}$. Let $a = (a_1, a_2, ta_1^2 - a_2^2)$ and $b = (b_1, b_2, tb_1^2 - b_2^2)$ be any two distinct points on our paraboloid. We will show that no other point on the line through a and b can lie on the paraboloid. If $ra + (1-r)b \in P_3$, then

$$t(ra_1 + (1-r)b_1)^2 - (ra_2 + (1-r)b_2)^2 = r(ta_1^2 - a_2^2) + (1-r)(tb_1^2 - b_2^2).$$

Distributing and moving terms around gives us

$$(r-1)rt a_1^2 - (r-1)ra_2^2 + 2(1-r)r(ta_1 b_1 - a_2 b_2) = (1-r)rt b_1^2 - (1-r)rb_2^2$$

and further manipulation lands us at

$$(1-r)r[t(a_1 - b_1)^2 - (a_2 - b_2)^2] = 0.$$

Since $a \neq b$ and t is a nonsquare, the bracketed expression must be nonzero, and thus $r = 0$ or 1 . That is, $ra + (1-r)b = a$ or b . Furthermore, the set $\{(x, 0, tx^2 - 0^2) : x \in \mathbb{F}_q\}$ cannot contain three collinear points, which implies that $\{(x, tx^2) : x \in \mathbb{F}_q\} \subset \mathbb{F}_q^2$ cannot either.

Now we can prove the result for P_d for $d \geq 3$. For S in the power set of $\{1, \dots, d\}$, define π_S to be the projection from \mathbb{F}_q^d onto the coordinates given by S . Suppose that P_d contains three distinct collinear points a, b , and c . Then there is some $s \in \{1, \dots, d\}$ so that $\pi_{\{s\}}(a)$, $\pi_{\{s\}}(b)$, and $\pi_{\{s\}}(c)$ are distinct. Notice however, that a projection of a line is either a line or a single point. We have already determined that $\pi_{\{1,2,3\}}(a)$, $\pi_{\{1,2,3\}}(b)$, and $\pi_{\{1,2,3\}}(c)$ cannot be distinct (as then $P_3 = \pi_{\{1,2,3\}}(P_d)$ would have three points on a line), and therefore

$$\pi_{\{1,2,3\}}(a) = \pi_{\{1,2,3\}}(b) = \pi_{\{1,2,3\}}(c).$$

For the same reason,

$$\pi_{\{3n-2, 3n-1, 3n\}}(a) = \pi_{\{3n-2, 3n-1, 3n\}}(b) = \pi_{\{3n-2, 3n-1, 3n\}}(c) \text{ for } 1 \leq n \leq \frac{d}{3},$$

and

$$\pi_{\{d-1, d\}}(a) = \pi_{\{d-1, d\}}(b) = \pi_{\{d-1, d\}}(c) \text{ when } d \equiv 2 \pmod{3}.$$

It follows that $a = b = c$, which is a contradiction.

References

- [1] Bateman, M. and N. H. Katz, *New bounds on cap sets*, J. Amer. Math. Soc. **25** (2012), no. 2, 585–613. [12](#)
- [2] R. Beheshti, *Lines on projective hypersurfaces*, J. Reine Angew. Math. **592** (2006), 121. [13](#)
- [3] P. Erdős, Z. Füredi, *The greatest angle among n points in the d -dimensional Euclidean space*, Ann. Discrete Math. **17** (1983), 275–283. [3](#)
- [4] V. Harangi, T. Keleti, G. Kiss, P. Maga, A. Máthé, P. Mattila, B. Strenner, *How large dimension guarantees a given angle?*, Monatsh. Math. **171** (2013), no. 2, 169–187. [2](#), [3](#)
- [5] D. Hart, A. Iosevich, D. Koh and M. Rudnev, *Averages over hyperplanes, sum-product theory in finite fields, and the Erdős-Falconer distance conjecture*, Trans. Amer. Math. Soc. **363** (2011), no. 6, 3255–3275. [5](#), [10](#)
- [6] D. Mubayi, J. Williford, *On the independence number of the Erdős-Renyi and projective norm graphs and a related hypergraph*, J. Graph Theory **56** (2007), no. 2, 113–127. [6](#)
- [7] T. Nagell, *Introduction to Number Theory*. New York: Wiley (1951). [8](#)
- [8] L. Vinh, *The number of occurrences of a fixed spread among n directions in vector spaces over finite fields*, Graphs Combin. **29** (2008), no. 6, 1943–1949. [10](#)

DEPARTMENT OF MATHEMATICS, ROCHESTER INSTITUTE OF TECHNOLOGY, ROCHESTER, NY 14623
E-mail address: mbbsma@rit.edu